Flight Readiness Technology Assessment
NASA EEE Parts Program

Assessment of DRAM Reliability from Retention Time Measurements

Udo Lieneweg
Duc N. Nguyen
Brent R. Blaes

Jet Propulsion Laboratory
California Institute of Technology

June 1998

Content

1. Summary 3
2. Introduction 3
3. Early Failures 4
4. Temperature Dependence 8
5. Stress Experiments 12
6. Later Failures 16
7. SEM Analysis 17
8. Conclusions 17
9. References 18
10. Acknowledgments 19
11. List of Figures and Tables 20

1. Summary

It was attempted to assess the reliability of DRAMs at the package level. Since the most critical function of a DRAM storage cell is to retain the charge which represents its state, data retention was measured at nominal conditions as a function of refresh time on several 5-V, 4-Mbit memories and one 3.3-V, 16-Mbit memory. Model distributions were fitted to the failure distributions and extrapolated to the time to first failure t_l. These times t_l are an indication of the maturity of each product. The ratio t_l/t_r, where t_r is the maximum specified refresh time, is a figure of merit for tolerance to degradation of the product. In the second phase of the project, the retention time distributions were measured at temperatures down to -30 °C and activation energies derived. The memories were then stressed at elevated voltages and temperatures for times up to 18 hours while they were continually refreshed at normal rate. After each stress a memory was subjected to a checkerboard test at nominal conditions. Then the retention times were remeasured at normal voltage and low temperatures. The hypothesis for this experiment, based on some literature reports, was that the leakage of the dielectric of the storage capacitors would be increased by charge injection under stress before breakdown of the dielectric. Low temperature was applied during measurement to separate this effect, which is based on tunneling through the dielectric, from leakage to the substrate and through the pass transistor, which is based on thermal carrier generation. The results are as follows: One type of 4-Mbit memory failed the checkerboard test in a regular pattern before any degradation of retention times could be detected. The other type, by Micron Technology, known to have internal reduction of the external supply voltage, showed a small degradation effect at early failures after stress at 10 V and 70 °C. Experiments were then started at the Micron Technology 16-Mbit product, which has no internal reduction of the 3.3-V supply voltage. No degradation was detected for stresses up to 8 V at 70 °C although a higher internal voltage was reached as in the 5-V product. It would be interesting to continue the experiment on the 16-Mbit device to higher stress voltages and compare the results to conventional reliability data published by the manufacturer for this device. In a side effort, the question was investigated on the Micron Technology 4-Mbit device, whether blocks of cells could be isolated and contacted, e.g., by FIB, to conduct direct stress experiments. The answer is probably yes, but with too much effort for the present task.

2. Introduction

Dynamic Random-Access Memories (DRAMs) store bits of information as charge of capacitors. Storage capacitors are formed between individual plates, which are connected each to one pass transistor, and a common cell plate. Leakage of the stored charge requires periodic refreshment, i.e., reading and writing back of the stored information before it degrades. The maximum time for which the information of a cell can still be correctly restored is called its retention time. Obviously the smallest retention time of all memory cells must be larger than the refresh time.

In older designs, the cell plate is at ground potential, and the capacitor is either charged to the supply voltage V_{cc}, representing a physical One state, or discharged to zero voltage, representing a physical Zero state. However, the differential nature of the sense amplifiers, which detect the charge states of the cells, requires that only one half of all cells represent logical states directly by physical states and that the other half represent logical states by complementary physical states. Consequently, if the memory is written full with (logical) Ones, the state of the half of the memory cells which is charged is subject to degradation with time due to charge leaking from the capacitors. Three leakage paths have been identified: first, subthreshold leakage through the pass transistor; second, leakage from the storage node of the transistor to the substrate; and third, leakage through the dielectric of the storage capacitor. Newer designs may bias the cell plate at $V_{cc}/2$ in order to reduce the electric field in the thinner dielectric of the storage capacitor [1]. After passing of its retention time, a charged cell has lost a certain threshold charge such that the remaining charge is detected as zero. This fixed threshold charge equals obviously the average leakage current times the retention time. Hence, the retention time is inversely proportional to the average leakage current, and we can measure the distribution of cell leakage currents by measuring the distribution of retention times.

It can be argued that the minimum retention time is a measure of the maturity of a DRAM product and that the ratio of the minimum retention time to the maximum specified refresh time is a figure of merit for tolerance to degradation of the product. In the next section we will characterize the early failures of different DRAMs and derive the figures of merit. We will then discuss the full failure distributions and their temperature dependence.

We will see that these retention failure distributions are very similar to the time-to-failure distributions of the breakdown of thin dielectrics. Therefore, we hypothesized that dielectric leakage may be a precursor to breakdown and that it may be increased through electrical and thermal stress before breakdown or other loss of function occurs. We report on experiments looking for such an effect, which would show up as a shift in the retention time distribution measured after the stress.

Finally, the question was investigated on one device, whether blocks of cells could be isolated and contacted, e.g., by Focused Ion Beam (FIB) to conduct direct conventional dielectric breakdown experiments.

3. Early Failures

Early retention failures for times up to t = 400 s were investigated with an Advantest T3342 tester on three types of 5-V, 4-Mbit DRAMs. Some characteristic data of these parts, supplied by the manufacturers, are listed in Table 1. All parts are packaged in SOJ plastic packages. After an initial checkerboard test, the memories were written with all Ones. Automatic refresh was disabled. After time steps which doubled each time, the contents were read back and the number of failed words counted. All this was done at nominal conditions, i.e., $V_{cc} = 5$ V and $T = 30$ °C.

Table 1: DRAM synopsis

Manufacturer	Hitachi	Texas Instruments	Micron Technology	Micron Technology
Part Number	HM514100C-S6	TMS44100-60DJ	MT4C4001JDJ-6	MT4LC4M4B1DJ6
Revision		E		
Supply Voltage	5 V	5 V	5 V	3.3 V
Words x Bits	4M x 1	4M x 1	1M x 4	4M x 4
Blocks x Bits	16 x 256K	32 x 128K	16 x 256K	64 x 256K
Process	CMOS	EPIC CMOS	CMOS Si-gate	
Min. Feature	0.8 um	0.5 um	0.3 um	0.3 um
CMOS Well		Twin		
Capacitor Type		Trench	Stack	Stack
Bit Capacitance		58 fF		
Power, active	605 mW		225 mW	530 mW
Power, standby	11 mW		3 mW	
Therm. Resist. ϑ_{jc}		2.89 °C/W	7 °C/W	3 °C/W
Max. Refresh Time	16 ms	16 ms	16 ms	128 ms

Temperature-controlled air was blown onto the package/wiring board assembly. The temperature was measured at this time with a thermocouple stuck between the package and the board.

The failure counts were converted into bit-failure rates F by dividing by one half of the bit capacity, i.e., 2,097,152 = 2M[1]. Failure times were normalized to the specified maximum refresh time of 16 ms. The data were first plotted on log-normal probability paper and then on log-extreme-value paper. Only the latter plots are presented here, Figs. 1 to 3, because

---

[1] This procedure is correct in a strict sense only for the 4M x 1 devices. For the 1M x 4 device it is correct as long as it can be assumed that only one bit per word fails. This should be a good assumption for the early failures since we have not observed any clustering, cf. below.

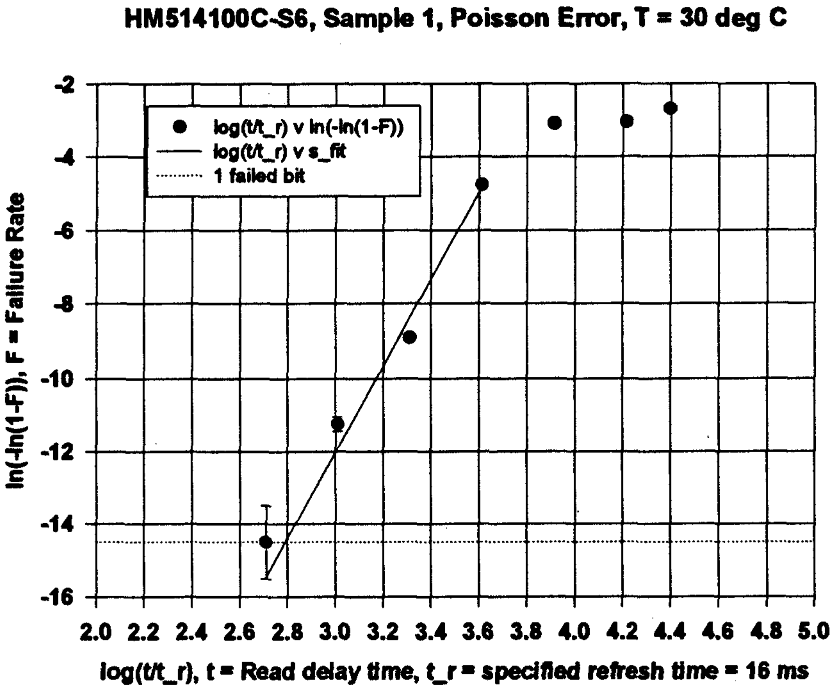

HM514100C-S6, Sample 1, Poisson Error, T = 30 deg C

Fig. 1: Failure time distribution of Hitachi 4M x 1 sample.
(Three upper points invalid due to saturation of counter.)

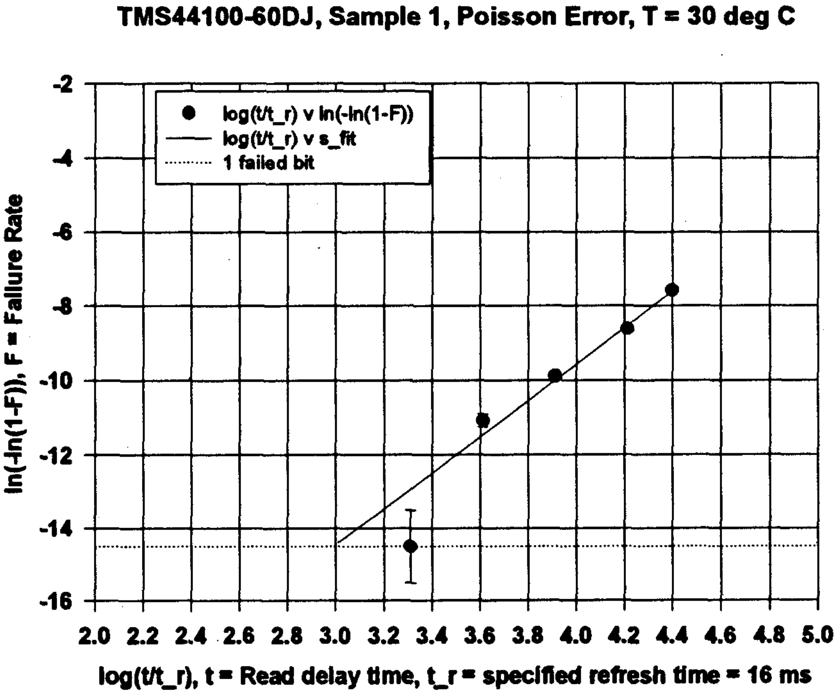

TMS44100-60DJ, Sample 1, Poisson Error, T = 30 deg C

Fig. 2: Failure time distribution of Texas Instruments 4M x 1 sample

of the better fit to a straight line, which would confirm that type of distribution. The points were plotted with error bars corresponding to a Poisson error equal to the square-root of counts. Lines were fitted to the points with the errors as weights. Variances from run to run and from sample to sample were explored on the Micron Technology 1M x 4 product. As Fig. 3 shows, the data reproduce within the Poisson error. The abscissa of the cross-over of each fitted line with the horizontal line indicating one failed bit was used to determine the representative time to one failure. This time t_1 and its ratio C_1 to the maximum specified refresh time are listed in Table 2.

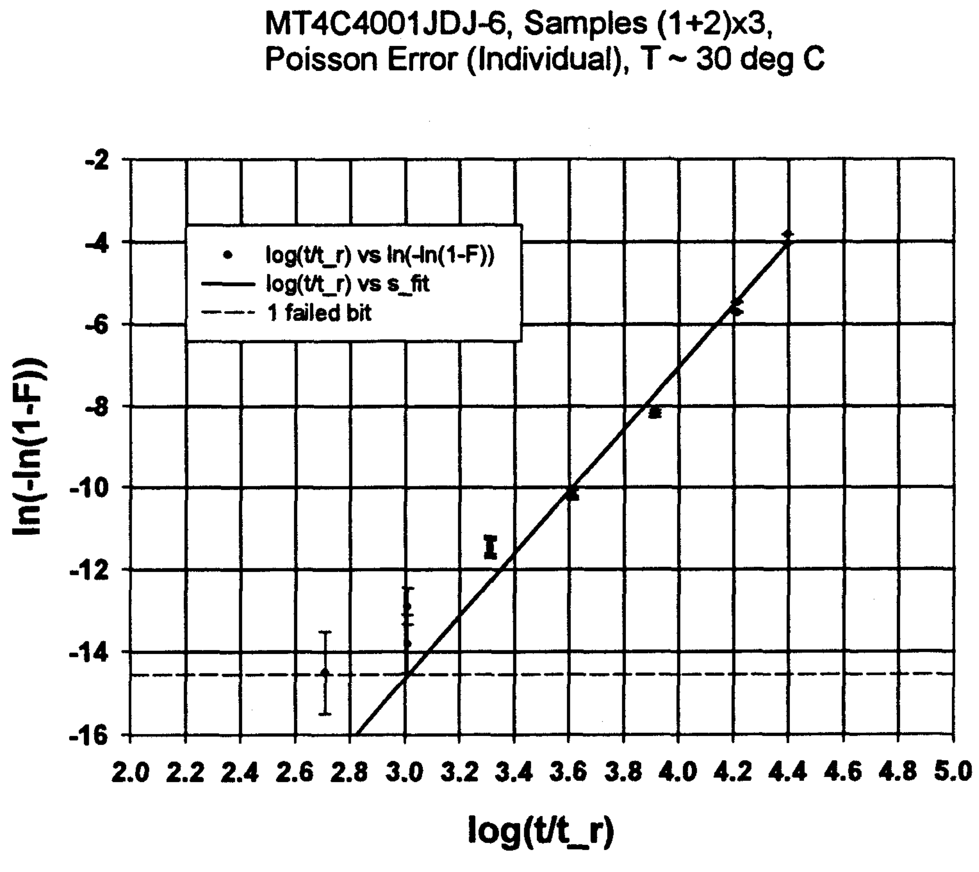


Fig. 3: Bit-failure time distributions of Micron Technology 1M x 4 Samples 1 and 2, each tested 3 times at 30 °C.

Table 2: Figures of merit of DRAMs for operation at 30 °C

Device	HI 4M x 1	TI 4M x 1	MT 1M x 4	MT 4M x 4
Max. Refresh Time t_r	16 ms	16 ms	16 ms	128 ms
Time to first failure t_1	10 s	16 s	16 s	9 s
$C_1 = t_1 / t_r$	630	1000	1000	70

It can be seen from the table that all 4-Mbit products have a large failure margin at 30 °C. The Micron Technology (MT) and the Texas Instrument (TI) devices are equally good, whereas the Hitachi (HI) device, which uses the oldest technology, is slightly inferior.

In a separate experiment, we convinced ourselves that the early failures were indeed of a random nature by recording their location in a certain block. For that purpose, the address space of the memory was divided into blocks of 128 x 256 = 32,768 bits. Visual inspection of the failure maps revealed no pattern or clustering.

4. Temperature Dependence

As mentioned in the introduction, there are three paths for storage capacitor charge to leak out. Besides through the capacitor dielectric, charge may leak through the substrate and the transistor channel. The two latter effects on the time to first-bit failure can be magnified by either increasing or decreasing the substrate bias as shown by Shaw et al. [2]. For an n-channel pass transistor, negative substrate biasing decreases the subthreshold current exponentially. However, at a too large negative bias, substantial current may be generated in the depletion region of the storage node's p-n junction. This current may be generated by thermal activation of electrons through near mid-gap centers and is proportional to the depletion width. At temperatures high enough to overcome the full bandgap of silicon, diffusion of minority carriers may also play a role. We have verified on the MT device that a small negative substrate bias is generated on-chip, which in effect suppresses both the subthreshold and the substrate current. In fact, the data in Ref. [2], taken on an MT 16-Mbit device, show a broad bias range from -0.25 V to -2.25 V with leakage larger than provided by both other mechanisms. It is conceivable that in this region dielectric leakage dominates.

In order to obtain further clues about the dominant leakage mechanism, we measured the retention time distributions at different temperatures. Because of the expected high activation energies of the competing mechanisms, we took measurements at lower temperatures, i.e., at 0 °C and -30 °C. The temperature measurement was improved by using a finer, 36-gauge, e-type thermocouple wire and attaching it with heat resistant, plastic adhesive tape and a drop of silicone heat-conduction paste to the top of the DRAM package. With this technique temperature readings were reproducible within 0.2 °C. From the measured package ("case") temperature T_c, the dissipated power P, and the published thermal resistance ϑ_{jc}, see Table 1, the chip ("junction") temperature T_j can be estimated from

$$T_j - T_c = \vartheta_{jc} P.$$

For the MT 1M x 4, the device with the highest published ϑ_{jc}, we measured $P = 5$ V × 16 mA = 80 mW under a 16-ms read cycle, giving a temperature difference between chip and $T_j - T_c = 7$ ° C/ W × 0.08 W = 0.6 °C. In the following we neglect this small difference.

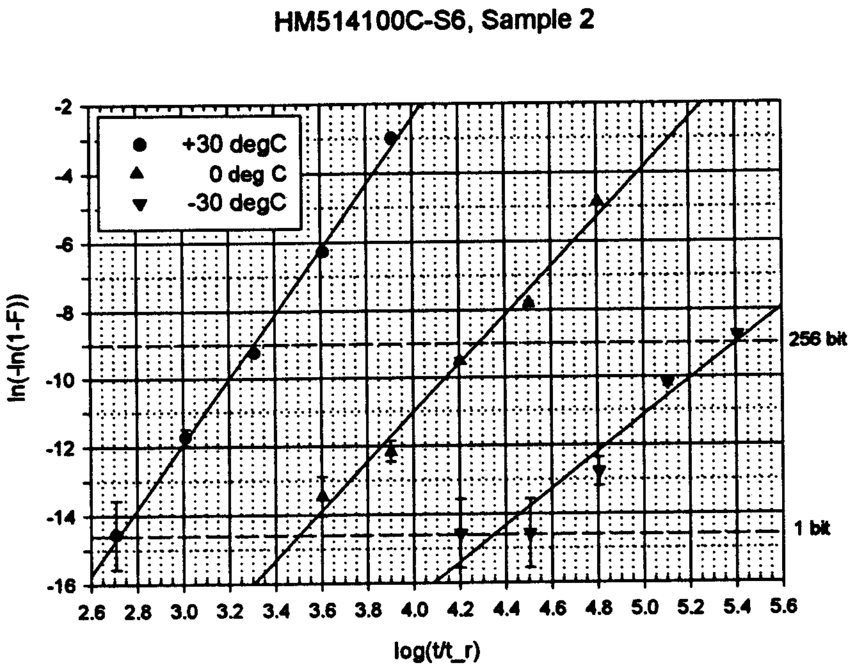


Fig. 4: Failure time distribution of Hitachi 4M x 1 sample at low temperatures

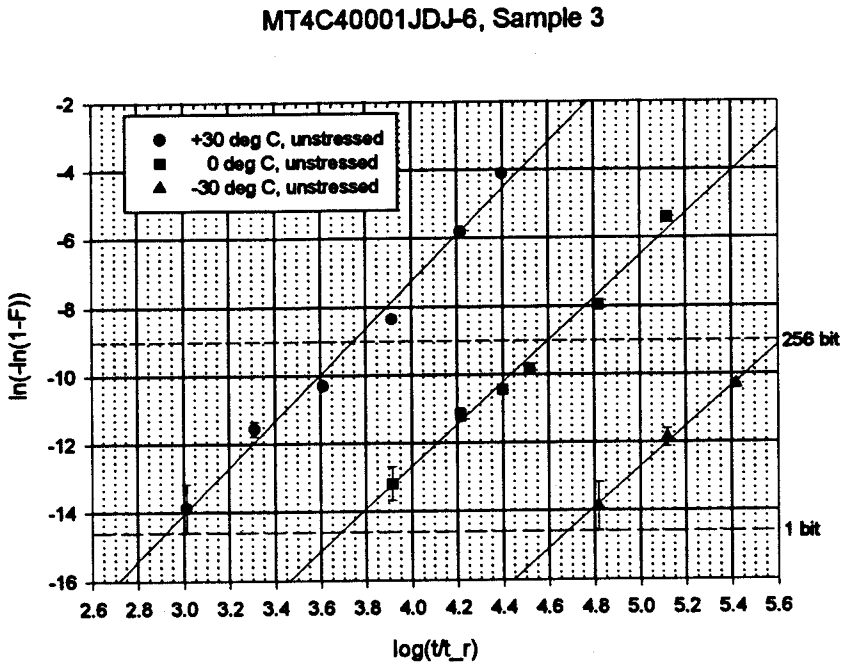


Fig. 5: Bit-failure time distribution of Micron Technology 1M x 4 sample at low temperatures

Figures 4 and 5 show the early failure distributions as a function of cooling for HI 4M x1 and the MT 1 x 4 DRAMs. The TI 4M x 1 product was dropped from further investigation because its sale was discontinued. The shift of the distributions at two bit levels with temperature was explored in Arrhenius plots, Fig. 6.

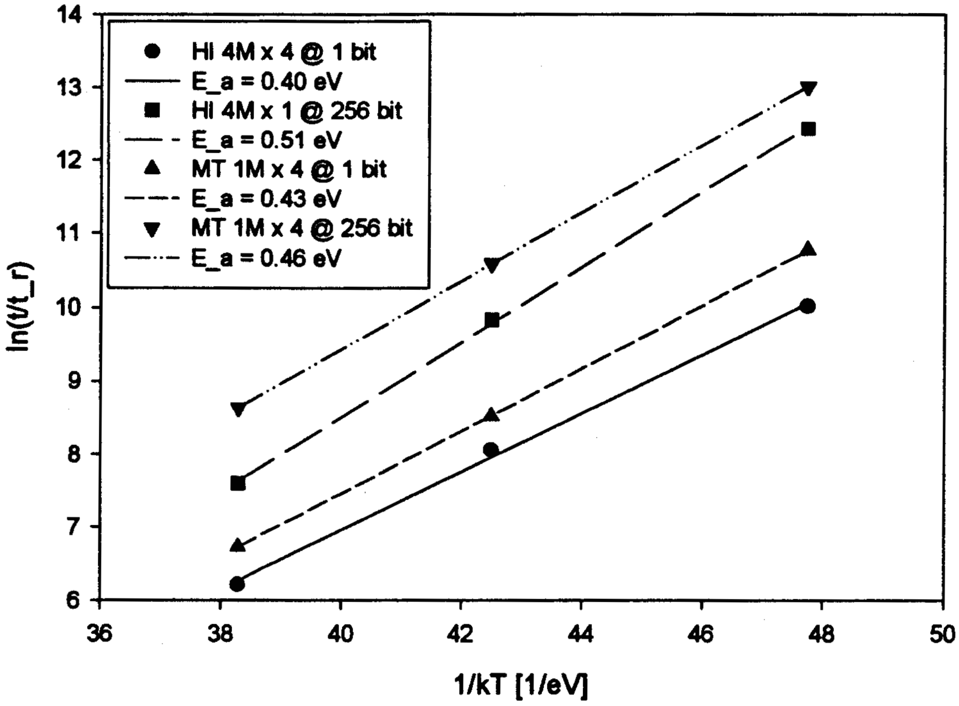


Fig. 6: Arrhenius plots of retention times of HI and MT 4-Mbit samples at two bit-levels

The plots prove that the shifts in retention times with temperature are thermally activated according to

$$t = t_n \exp(E_a/kT),$$

where $k = 8.62 \times 10^{-5}$ eV/K is Boltzmann's constant, t_n a parameter depending on the bit-level, and E_a activation energies listed in the legend. The HI and the MT samples have activation energies of 0.40 eV and 0.43 eV, respectively, at the 1-bit level and somewhat larger energies at the 256-bit level. At least the very early failures have activation energies which are distinctly lower than half the bandgap of silicon, which is 1.12/2 eV = 0.56 eV at 300 K, so that we may indeed be looking at an effect different from substrate leakage.

5. Stress Experiments

It is compelling that the distributions of early retention times are of the same type as those of times-to-dielectric breakdown, namely extreme-value distributions. This led us to the hypothesis that we were looking at dielectric leakage which is a precursor to breakdown. We further hypothesized that it may be possible to subject the storage dielectric to electric and thermal stress so that the leakage increases, but the memory is still testable.

Increase in low-level current in thin oxides after voltage stress has indeed been reported by a number of authors [3], [4], [5]. They agree in the interpretation that the stress generates traps in the oxide, which act as tunneling centers. The latest of these papers [5] found that the increase in the leakage current is proportional to the trap density. Trap energies of about 0.3 eV below the conduction band were representative. The breakdown occurring after higher stress was related to a critical density of the generated traps.

We performed stress experiments with the DRAMs by operating them at increased supply voltages and 70 °C for 17-18 h at a time. During that stress time the memory was written with all Ones and read back periodically at the specified refresh time (with the levels of the address signals increased accordingly). It was monitored that no cell flipped during that time. After each stress, the devices were tested with a checkerboard pattern at 5 V and 30 °C and than the retention time distributions were remeasured at 5 V and +30 °C, 0 °C, and -30 °C.

Of the stressed HI 4M x 1 samples, Sample 1 showed no stress effect for voltages stepped up to 12 V in 1-V steps, but failed the checkerboard test after stress at 13 V. Failure mapping showed a regular pattern related to certain address bits, so that malfunction of decoding can be suspected. Sample 3 was then stressed at 12.5 V. No significant stress effect can be seen in Fig. 7.

The sample was also used to explore the effect of stress to later failures. Since no late failures were recorded on the unstressed sample, it was compared to a distribution from the unstressed Sample 4 in Fig. 8. No significant changes can be seen at the later failures, either. (The deviation of the later failures from the extreme-value distribution will be discussed below.)

Next we stressed MT 1M x 4 samples. The results for stresses up to 10 V are shown in Fig. 9. Note that the word-failure rate F_w = counts/524,288 is now used. Here, indeed, we see a small but significant shift after stress at both compared low temperatures.

A publication by MT [6] indicates that their 5-V, 1M x 4 product has an internal reduction of the external supply voltage to about 3.3 V, which is effective up to an external voltage of about V_{cc} = 6.5 V. Beyond that the internal voltage rises almost as fast as the external one, reaching 6.5 V at V_{cc} = 10 V. At this point, we decided to discontinue the stress of this product and, instead continue with an MT 3.3-V, 4M x 4 product. This device, while having the same cell layout, has no internal reduction of the supply voltage [7]. The larger capacity should also expose more early failures. General specifications of the 16-Mbit DRAM have been added to Table 1.

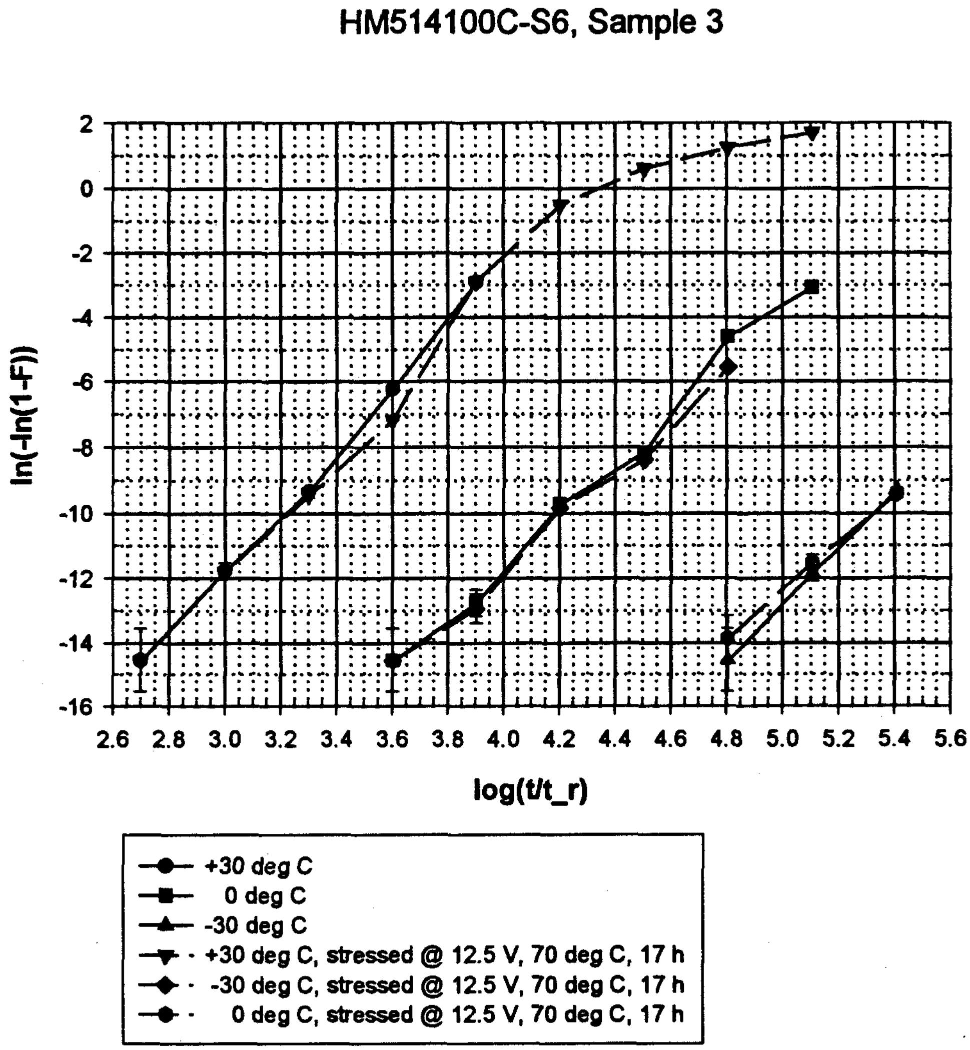


Fig. 7: Retention time distributions for HI 4M x 1 sample before and after highest stress which was survived with functionality

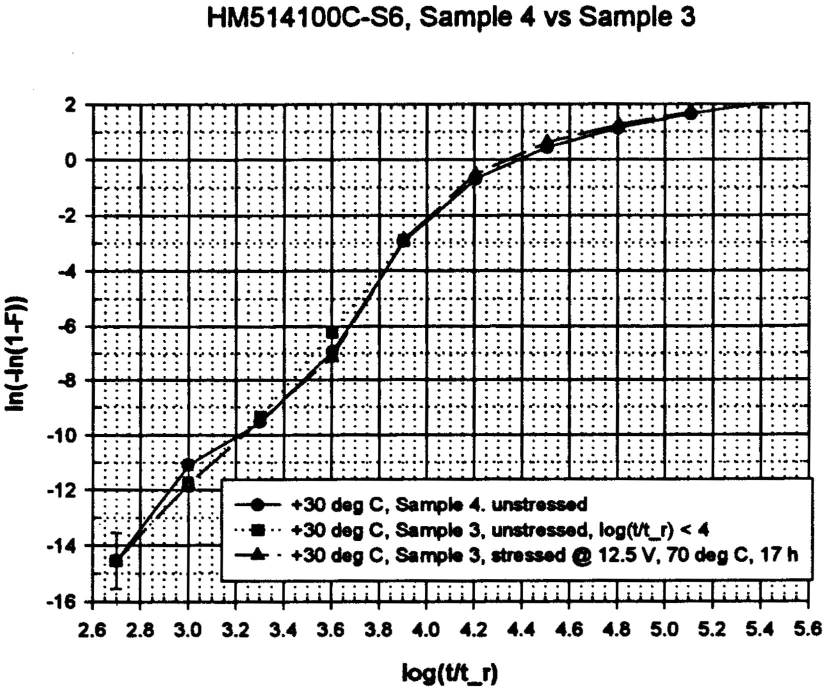


Fig. 8: Comparison of total distributions of stressed and unstressed HI samples

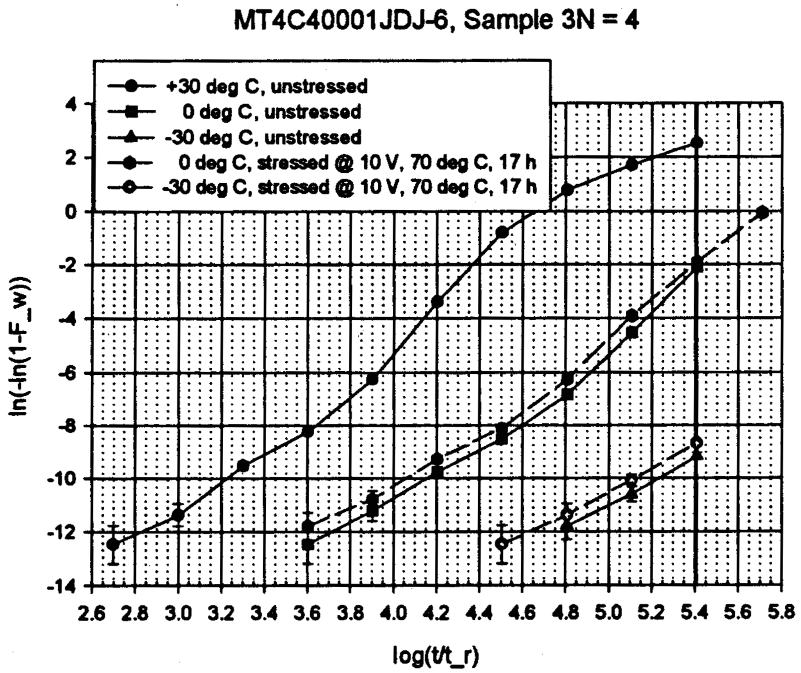


Fig. 9: Stress effect on retention time distribution of MT 1M x 4 sample.

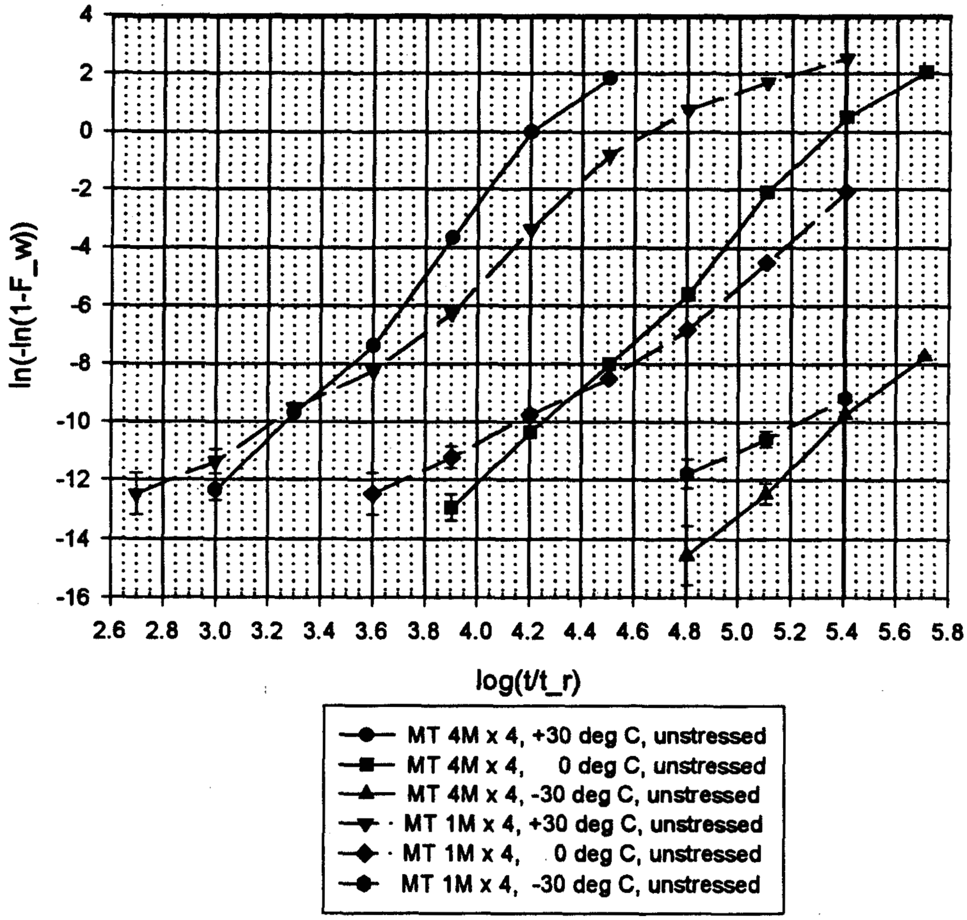


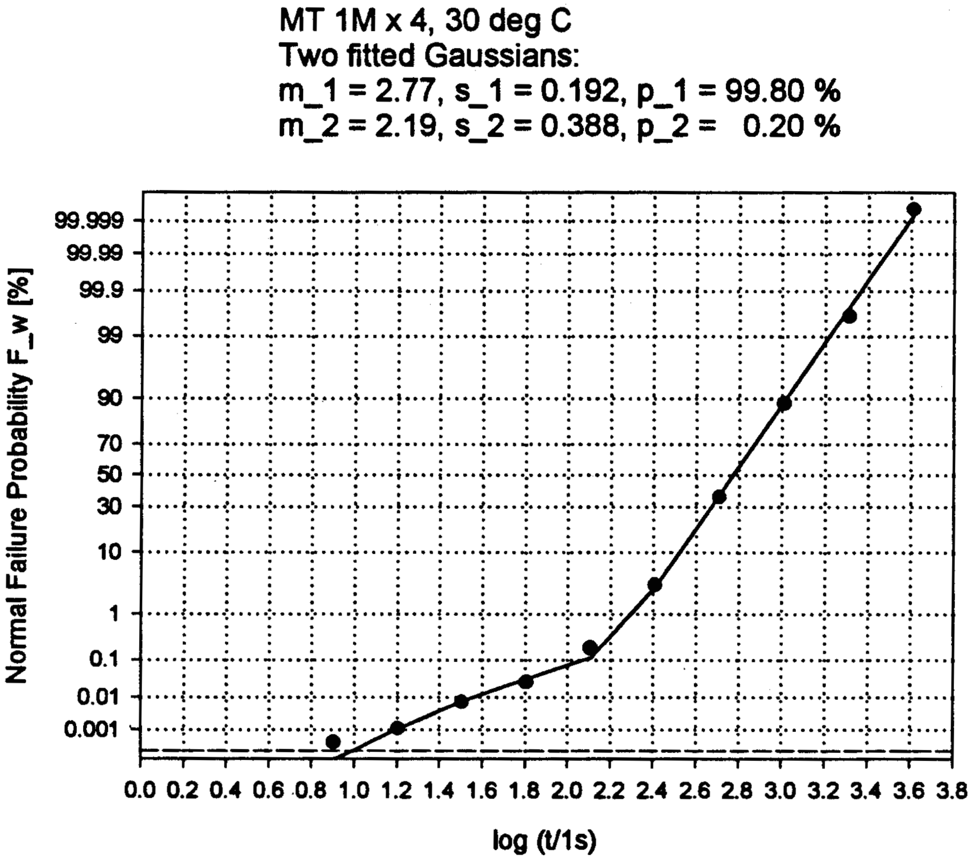
Fig. 10: Comparison of distributions from MT 4M x 4 DRAM and 1M x 4 DRAM. (For absolute comparison times are normalized to 16-ms refresh time in both cases although the 4M x 4 allows 128 ms.)

Fig. 10 shows the retention time distributions for the 4M x 4 DRAM compared to the ones of the 1M x 4. In order to compare scales directly times were normalized to a 16-ms refresh time although the 4M x 4 allows 128 ms. It is interesting to note that the distributions are not identical in spite of the same cell layout. Thus, some other changes must have been implemented, e.g., in the process. As the figure shows, improvements have been achieved "only" at the early failures; the later failures have actually deteriorated as the distributions became steeper. Extrapolation of the 30 °C-data to 1-bit failure gives the figures of merit added to Table 2. Note that now $C_l = t_l / 128$ ms has shrunk to a mere 70. Especially high-temperature operation of the 16-Mbit device may become unreliable.

Stressing the 4M x 4 at 7 V and 70 °C caused no change of the distributions measured at 0 °C and –30 °C. A following stress at 8 V showed also no effect on the distribution at 0 °C. Further measurements were not made due to closure of the task.

6. Later Failures

We return now to the observation that the distribution plots on extreme-value paper became curved for later failures. We therefore replotted some curves on log-normal paper, which expands the late failures. Fig. 11 shows the replotted 30 °C-distribution of MT 1M x 4 Sample 3N = 4, which was previously plotted in Fig. 9.



MT 1M x 4, 30 deg C
Two fitted Gaussians:
m_1 = 2.77, s_1 = 0.192, p_1 = 99.80 %
m_2 = 2.19, s_2 = 0.388, p_2 = 0.20 %

Fig. 11: Failure distribution at 30 °C of sample from Fig. 9 replotted on normal probability paper and fitted with two Gaussian distributions. Dashed line: 1-bit failure

While we obtain now a very good fit with a normal distribution for failures above 1%, the early failures appear now as a distinctive separate mode. For a best fit to two Gaussians *on the normal scale,* the parameters indicated in the figure were found. To obtain such a fit, the tails of the distribution had to be weighted heavily. We used $F_w^{-a} + (1-F_w)^{-a}$ as weights to F_w with $a = 1.75$. Note that the second mode contributes only 0.2 %. A fit of the second mode with an extreme-value distribution is likely to still improve the fit. A small third mode (< 1%) of very late failures with counts > 524,288 = 0.5M was cut off due to the normalization of the counts to that number. These failures stem likely from cells with complementary charge, which may flip with a very small probability. As late failures, however, they do not have remotely the importance of the early ones.

7. SEM Analysis

The plastic packages of 4-Mbit parts were dissolved and optical micrographs taken of the chips. Scanning Electron Microscopy (SEM) analysis of 4-Mbit Micron Technology parts was performed (a) after stripping of overglass and (b) after stripping of metal interconnects. Memory cell construction (stacked storage capacitors) and memory blocks were identified. We analyzed connections to the cell plate and determined that the cell plate could be contacted by making contact to a metal line at the periphery of each block, but that six polysilicon lines per block would have to be cut to disconnect the cell plate from other circuitry. (Contemporary technology biases the cell plate at one half of the supply voltage.) We found also that the p-substrate contact is connected to the lead frame. A voltage could then be applied between substrate (positive) and cell plate (negative) to directly measure storage dielectric breakdown. The measurement could be made with this polarity only, since the p-n diode comprised of the substrate and the storage node implant must be forward biased. Cross-sections published by Micron Technology show, however, that the storage node plates of the storage capacitors are corrugated in contrast to the smooth cell plate. This would favor electron injection from the storage node side, a condition which could not be realized with the outlined approach.

While it may be possible to make all these cuts and contacts by Focused Ion Beam (FIB) techniques, it would be very difficult and certainly beyond the means of this task. A major concern was also with redeposition of conducting material at the capacitor edges and damage to the dielectric from ions.

8. Conclusions

Times to failure as a function of time between writing and reading all Ones (retention times) were measured on several 4-Mbit DRAMs and one 16-Mbit device. The main failure mode is clearly log-normally distributed. A small mode of early failures is best fit by an extreme-value distribution. Similar two-mode distributions are often described in dielectric breakdown with the early mode being attributed to “defects”. We found some evidence that dielectric leakage is a precursor to the actual breakdown, although more data are needed for a solid conclusion.

These findings are of great importance for fast reliability assessment procedures on a finished DRAM product. We have demonstrated a statistically superior technique for time to first failure estimation. If the deterioration of the retention times with stress can be confirmed in further experiments, this technique can substitute for time consuming conventional failure experiments or measurements on special breakdown test structures.

9. References

[1] P. Mazumder and K. Chakraborty, "Testing and testable design of high-density random-access memories", Kluwer Academic Publ., Boston 1996

[2] D.C. Shaw, G.M. Swift, D.J. Padgett, and A.H. Johnston, "Radiation Effects in Five Volt Advanced Lower Voltage DRAMs", IEEE Trans. Nucl. Sc. 41, 2452 (1994)

[3] P. Olivio, T.N. Nguyen, and B. Ricco, " High-Field-Induced Degradation in Ultra-Thin SiO_2 Films", IEEE Trans. Electr. Dev. 35, 2259 (1988)

[4] R. Moazzami and C. Hu, "Stress-Induced Current in Thin Silicon Dioxide Films", IEEE Proc. IEDM 92, 130 (1992)

[5] D.J. Dumin, J.R. Maddux, R.S. Scott, and R. Subramoniam, "A Model Relating Wearout to Breakdown in Thin Oxides", IEEE Trans. Electr. Dev. 41, 1570 (1994)

[6] Micron Technology, "DRAM 4 Meg Reliability Monitor", Rev. D37M 8/96, Boise, Idaho 1996

[7] Micron Technology, "DRAM 16 Meg Reliability Monitor", Rev. D42S/T 3/97, Boise, Idaho 1997

10. Acknowledgments

The research described in this report was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the National Aeronautics and Space Administration's Electrical, Electronic and Electromechanical Parts Program. Sammy Kayali is the manager of that program at JPL.

We wish to thank Duc Vu, Jim Okuno, and Ken Evans of the Failure Analysis Lab for the structural analysis. Frank Stott and Donna Turnbow were very helpful in the procurement of the parts. Valuable discussions with Alan Johnston and Russ Lawton; and technical support by Mike O'Connor and Otto VonWachter are acknowledged.

11. List of Figures and Tables

Fig. 1: Failure time distribution of Hitachi 4M x 1 sample

Fig. 2: Failure time distribution of Texas Instruments 4M x 1 sample

Fig. 3: Bit-failure time distributions of Micron Technology 1M x 4 Samples 1 and 2, each tested 3 times at 30 °C.

Fig. 4: Failure time distribution of Hitachi 4M x 1 sample at low temperatures

Fig. 5: Bit-failure time distribution of Micron Technology 1M x 4 sample at low temperatures

Fig. 6: Arrhenius plots of retention times of HI and MT 4-Mbit samples at two bit-levels

Fig. 7: Retention time distributions for HI 4M x 1 sample before and after highest stress which was survived with functionality

Fig. 8: Comparison of total distributions of stressed and unstressed HI samples

Fig. 9: Stress effect on retention time distribution of MT 1M x 4 sample

Fig. 10: Comparison of distributions from MT 4M x 4 DRAM and 1M x 4 DRAM.

Fig. 11: Failure distribution at 30 °C of sample from Fig. 9 replotted on normal probability paper and fitted with two Gaussian distributions. Dashed line: 1-bit failure

Table 1: DRAM synopsis

Table 2: Figures of merit of DRAMs for operation at 30 °C